



Vers une interprétation qualitative de comportements cinématiques dans la scène à partir du mouvement apparent

Edouard Francois, Patrick Bouthemy

► To cite this version:

Edouard Francois, Patrick Bouthemy. Vers une interprétation qualitative de comportements cinématiques dans la scène à partir du mouvement apparent. [Rapport de recherche] RR-1081, INRIA. 1989. inria-00075478

HAL Id: inria-00075478

<https://hal.inria.fr/inria-00075478>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNITÉ DE RECHERCHE
IRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 1081

Programme 6
Robotique, Image et Vision

VERS UNE INTERPRETATION QUALITATIVE DE COMPORTEMENTS CINEMATQUES DANS LA SCENE A PARTIR DU MOUVEMENT APPARENT

Edouard FRANCOIS
Patrick BOUTHEMY

Août 1989



VERS UNE INTERPRÉTATION QUALITATIVE DE COMPORTEMENTS CINÉMATIQUES DANS LA SCÈNE À PARTIR DU MOUVEMENT APPARENT

TOWARDS A QUALITATIVE INTERPRETATION OF OBJECT MOTION IN THE SCENE FROM OPTICAL FLOW

Publication Interne n° 481 - Juin 1989 - 40 Pages

Edouard FRANCOIS et Patrick BOUTHEMY
IRISA/Centre INRIA de Rennes
Campus de Beaulieu
35042 Rennes Cedex

RÉSUMÉ

Cette étude est une première approche vers une interprétation qualitative de comportements cinématiques dans la scène à partir du mouvement apparent dans la séquence d'images. La démarche adoptée se décompose en trois étapes. Premièrement, suite à un développement linéaire, on met en évidence des descripteurs adaptés décrivant bien le champ des vitesses apparentes. Dans un second temps, en passant par le torseur cinématique 3D attaché à un solide en mouvement, on établit le lien existant entre ces descripteurs et le mouvement 3D. Ceci permet la définition d'un jeu de symboles ou étiquettes correspondant à des configurations cinématiques données. La dernière étape est la validation des étiquettes, abordée selon une approche statistique, basée sur des critères d'information.

MOTS CLÉS

Séquence d'images, champ de vitesses, description symbolique, critères d'information, interprétation du mouvement.

ABSTRACT

This study is a first approach toward a qualitative interpretation of kinematic behaviour in the scene, from the apparent motion in an image sequence. The problem is addressed in the following way : from a linear development, descriptors that well describe the apparent velocity field are determined; using the 3D kinematic relation attached to a moving object, the relation between these descriptors and the 3D motion is established; with this link, we can define a set of labels corresponding to several kinematic configurations; the label validation step is then solved using a statistical approach, based on information criteria.

KEY WORDS

Image sequence, velocity field, symbolic description, information criteria, motion interpretation.

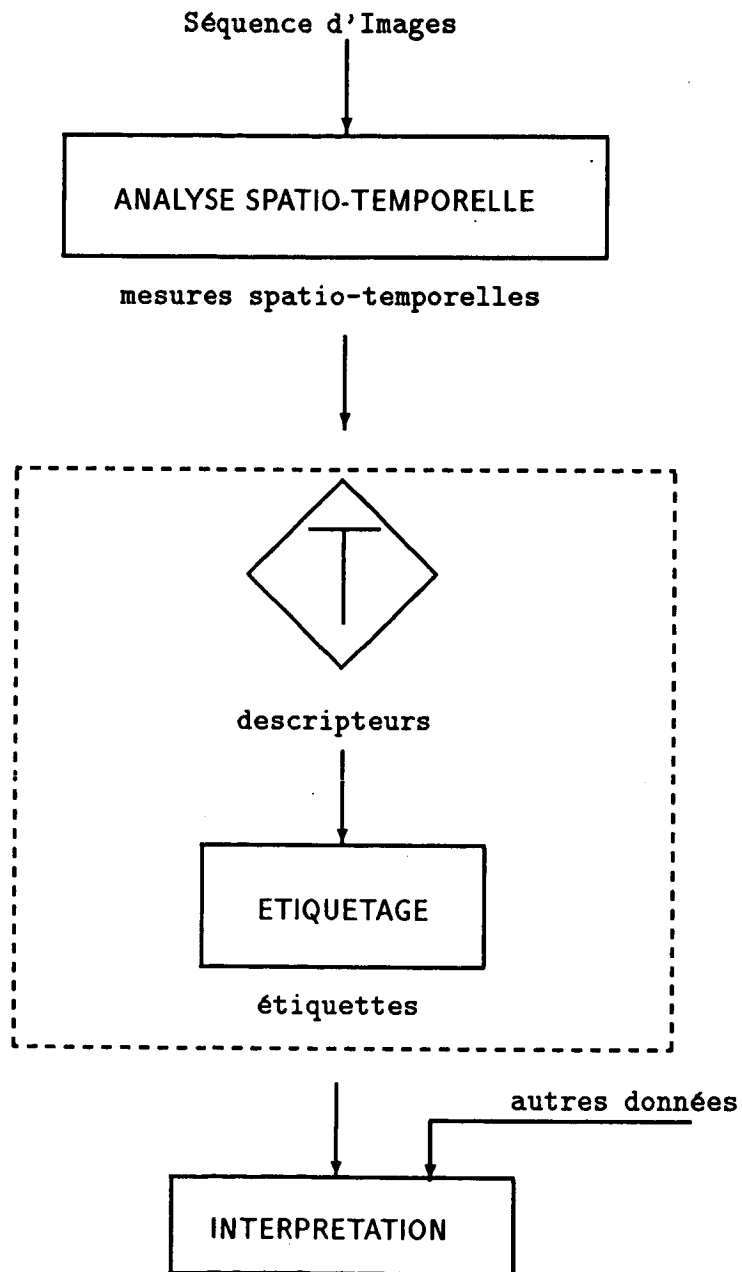
1 Introduction

Dans le domaine de l'analyse de scène dynamique, il est évident qu'existent des besoins de compréhension haut-niveau des comportements cinématiques observés dans la scène (par référence au capteur, ou dans un référentiel absolu). Il s'agit d'abord de besoins quantitatifs, c'est-à-dire de mesures explicites 3D+t. Mais on peut aussi exprimer des besoins qualitatifs dans des phases de décision, de diagnostic, de raisonnement ou autres. Ainsi des objectifs de surveillance de trafic, d'aide à la conduite automobile, d'aide au diagnostic médical pourront nécessiter une approche qualitative [8, 9, 16]. Si on se place dans ce domaine symbolique, il est possible de distinguer plusieurs classes d'information telles que la trajectographie, l'évolution dans le temps, l'analyse du type de mouvement. C'est à ce dernier point que l'on s'intéresse. Les motivations d'une telle étude sont diverses : l'interprétation qualitative de mouvement peut intervenir dans un contexte routier, afin de détecter des obstacles éventuels pour un "observateur", ce dernier étant mobile. Plus largement, elle peut être utile dans une démarche d'aide à la navigation, dans toute sorte de contexte.

Son but est de décrire de façon qualitative, à partir d'une séquence d'images, le comportement cinématique des objets de la scène (objets fixes, animés de mouvements frontaux ou transverses par rapport à l'axe de visée, rotation...). Il s'agit donc de reconnaître et d'étiqueter des zones dans l'image correspondant à des configurations cinématiques données dans la scène. La première étape est de déterminer un jeu de symboles qui permettent cette description. Le passage d'un contexte numérique (la séquence d'images, ou le champ des vitesses apparentes déduit d'une analyse spatio-temporelle) à un contexte symbolique suppose la mise en place de descripteurs en entrée du processus, dont la sortie est constituée d'un lot d'étiquettes ou symboles, décrivant le mouvement de façon qualitative. Par un développement linéaire du champ des vecteurs vitesse dans l'image et via certaines hypothèses, on peut mettre en évidence des descripteurs significatifs des différents mouvements 3D possibles des objets de la scène. Ces descripteurs permettent la définition d'un jeu d'étiquettes qualifiant différents comportements cinématiques. De plus, ils évitent de passer par une étape d'analyse 3D explicite. Une telle analyse implique en effet des problèmes de calibration, d'estimation de nombreux paramètres (paramètres explicites du mouvement 3D, profondeur...), que l'on préférerait éviter. Le choix des descripteurs et la définition des étiquettes sont traitées dans la Partie 2.

Chaque étiquette correspondant à un modèle particulier, l'étape de choix de l'étiquette revient à une étape de validation de modèle. Elle est abordée par une approche statistique basée sur des critères d'information (Partie 3). Les résultats et la conclusion sont présentés dans les Parties 4 et 5.

APPROCHE GENERALE



2 Choix des descripteurs. Définition des étiquettes

L'objectif global de l'étude est en quelque sorte d'effectuer de la reconnaissance de mouvement. Dans le domaine de la reconnaissance de forme, un point primordial est le choix des descripteurs. Ici aussi, cette étape est importante. Il s'agit de mettre en évidence des descripteurs adaptés à une interprétation 3D de la vitesse apparente dans l'image, projection du mouvement 3D dans la scène, et contenant des informations pertinentes et si possible indépendantes.

2.1 Autre formulation du modèle de vitesse linéaire

Considérons un point p de coordonnées (x_p, y_p) dans le plan image. Le développement au premier ordre du vecteur vitesse $\vec{\omega}$ autour de ce point s'écrit, en négligeant les termes d'ordre supérieur :

$$\vec{\omega}(x, y) = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_p \\ b_p \end{pmatrix} + \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} x - x_p \\ y - y_p \end{pmatrix} \quad (1)$$

$$\text{avec} \quad \alpha = \frac{\partial u}{\partial x}, \beta = \frac{\partial v}{\partial x}, \gamma = \frac{\partial u}{\partial y}, \delta = \frac{\partial v}{\partial y}$$

La matrice $M = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$ peut se réécrire de la façon suivante [13] :

$$M = \frac{1}{2}(\text{trace } M)I + \frac{1}{2}(M - M^T) + \frac{1}{2}[M + M^T - (\text{trace } M)I] \quad (2)$$

Cette nouvelle décomposition met en évidence des termes déterminants pour l'étiquetage :

- $\text{trace } M = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$ est la divergence de $\vec{\omega}(x, y)$ (que l'on notera *div*)

- $M - M^T = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

fait apparaître le terme $(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})$ correspondant au rotationnel de $\vec{\omega}(x, y)$ (noté *rot*)

- $M + M^T - (\text{trace } M)I = \begin{bmatrix} \delta - \alpha & \beta + \gamma \\ \beta + \gamma & \alpha - \delta \end{bmatrix}$ met à jour des termes "hyperboliques"

$$\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \text{ et } \left(\frac{\partial v}{\partial y} - \frac{\partial u}{\partial x} \right), \text{ qu'on note } hyp1 \text{ et } hyp2 .$$

Dans cette nouvelle base, la matrice M s'écrit donc :

$$M = \frac{1}{2} \text{div} [D] + \frac{1}{2} \text{rot} [R] + \frac{1}{2} \text{hyp1} [H1] + \frac{1}{2} \text{hyp2} [H2]$$

où

$$[D] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, [R] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, [H1] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, [H2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Cette décomposition semble particulièrement intéressante pour les raisons suivantes : tout d'abord, la relation (2) correspond à une décomposition de M dans une base orthogonale. De plus, les termes *div*, *rot*, *hyp1* et *hyp2* sont des descripteurs qui expriment l'allure d'un champ de vecteurs, chacun d'eux contenant une information caractéristique de ce champ. Ils sont informatifs pour tout champ de vecteurs, même lorsque le champ est plus complexe qu'un champ linéaire pur : lorsqu'on dispose par exemple d'un champ de type quadratique, l'interprétation peut se suffire de la seule partie linéaire. Enfin, deux descripteurs différents contiennent des informations différentes sur le champ.

2.2 Relations entre le mouvement 3D et le mouvement perçu dans l'image

Un autre point déterminant du choix de ces descripteurs vient de la signification et du lien physiques qu'ils présentent avec le mouvement 3D lorsque le champ analysé est un champ de vecteurs vitesse. Si l'on observe ces descripteurs pour différents mouvements, on aboutit aux conclusions suivantes [3, 4, 14, 15] : le terme de divergence met en évidence une translation le long de l'axe de visée, le terme de rotationnel un mouvement de rotation autour de cet axe, les termes hyperboliques traduisant des mouvements plus complexes.

Ceci peut se confirmer de manière explicite : pour cela, il suffit de considérer la projection sur le plan image du vecteur vitesse, dans l'espace 3D, \vec{V} d'un point d'un solide en mouvement, que l'on peut toujours représenter à l'aide du vecteur de translation instantanée \vec{T} , et du vecteur de rotation instantanée $\vec{\Omega}$, du torseur cinématique associé au solide, [7] (on peut noter que le rapport entre ces descripteurs et le mouvement 3D est mis en évidence d'une autre façon dans [4]). On part donc de la relation suivante:

$$\vec{V} = \vec{T} + \vec{\Omega} \wedge \vec{R}$$

où $\vec{R}^T = (X, Y, Z)$: vecteur de coordonnées du point dans l'espace
 $\vec{T}^T = (U, V, W)$
 $\vec{\Omega}^T = (A, B, C)$

$$\vec{v}^T = (\dot{X}, \dot{Y}, \dot{Z})$$

l'expression de chaque composante étant

$$\begin{cases} \dot{X} = U + BZ - CY \\ \dot{Y} = V - AZ + CX \\ \dot{Z} = W + AY - BX \end{cases} \quad (3)$$

Le repère (X, Y, Z) est choisi de telle sorte que (X, Y) soit un plan parallèle au plan image, l'origine du repère étant placée au centre de projection. La projection d'un point de l'espace (X, Y, Z) sur le plan image donne le point de coordonnées (x, y) dans ce plan, où $x = f.X/Z$ et $y = f.Y/Z$, f désignant la focale de la caméra. Pour simplifier les équations, on pose $f=1$, ainsi $x = X/Z$ et $y = Y/Z$.

On peut aisément retrouver les composantes du vecteur vitesse \vec{w} associé au mouvement du point (x, y) dans l'image :

$$\begin{cases} \dot{x} = u = \frac{1}{Z} \cdot \frac{\partial X}{\partial t} - \frac{X}{Z^2} \cdot \frac{\partial Z}{\partial t} \\ \dot{y} = v = \frac{1}{Z} \cdot \frac{\partial Y}{\partial t} - \frac{Y}{Z^2} \cdot \frac{\partial Z}{\partial t} \end{cases} \quad (4)$$

soit, après remplacement de \dot{X} , \dot{Y} et \dot{Z} par leurs expressions (relation (3)) :

$$\begin{cases} u = \frac{U}{Z} + B - Cy - x \frac{W}{Z} - Axy + Bx^2 \\ v = \frac{V}{Z} - A + Cx - y \frac{W}{Z} + Bxy - Ay^2 \end{cases} \quad (5)$$

Pour retrouver les termes du modèle linéaire, il faut introduire une notion supplémentaire, le voisinage du point P dans l'espace, qui exprimera la nature de l'élément de surface autour de P . Cette information est décrite par le développement en série de Taylor de la profondeur $Z = Z(X, Y)$:

$$Z = Z_0 + \left(\frac{\partial Z}{\partial X} \right)_0 \cdot X + \left(\frac{\partial Z}{\partial Y} \right)_0 \cdot Y + o^2$$

Le développement au premier ordre de $\frac{1}{Z}$ donne, après remplacement de $\frac{X}{Z}$ par x et $\frac{Y}{Z}$ par y :

$$\frac{1}{Z} = \frac{1 - \gamma_1 \cdot x - \gamma_2 \cdot y}{Z_0},$$

où $\gamma_1 = \left(\frac{\partial Z}{\partial X}\right)_0$ et $\gamma_2 = \left(\frac{\partial Z}{\partial Y}\right)_0$

que l'on réinjecte dans la relation (5) :

$$\left\{ \begin{array}{l} u = \left(\frac{U}{Z_0} + B\right) + \left(-\frac{\gamma_1}{Z_0}U - \frac{W}{Z_0}\right)x + \left(-\frac{\gamma_2}{Z_0}U - C\right)y \\ \quad + \left(\frac{\gamma_1}{Z_0}W + B\right)x^2 + \left(\frac{\gamma_2}{Z_0}W - A\right)xy \\ v = \left(\frac{V}{Z_0} - A\right) + \left(-\frac{\gamma_1}{Z_0}V + C\right)x + \left(-\frac{\gamma_2}{Z_0}V + \frac{W}{Z_0}\right)y \\ \quad + \left(\frac{\gamma_2}{Z_0}W - A\right)y^2 + \left(\frac{\gamma_1}{Z_0}W + B\right)xy \end{array} \right. \quad (6)$$

Ceci permet de relier les descripteurs *div*, *rot*, *hyp1* et *hyp2* aux paramètres du mouvement 3D :

$$\left\{ \begin{array}{l} div = -2\frac{W}{Z_0} - \gamma_1\frac{U}{Z_0} - \gamma_2\frac{V}{Z_0} \\ rot = 2C - \gamma_1\frac{V}{Z_0} + \gamma_2\frac{U}{Z_0} \\ hyp1 = -\gamma_1\frac{U}{Z_0} + \gamma_2\frac{V}{Z_0} \\ hyp2 = -\gamma_1\frac{V}{Z_0} - \gamma_2\frac{U}{Z_0} \end{array} \right. \quad (7)$$

2.3 Définition des étiquettes

Grâce à cette décomposition, il est désormais possible, à partir des descripteurs qu'elle introduit, de dresser un tableau décrivant les valeurs des descripteurs pour les différents mouvements 3D possibles.

On note :

- T_Z : translation le long de l'axe de visée
- T_{XY} : translation parallèle au plan image
- R_Z : rotation autour de l'axe de visée

- R_{XY} : rotation autour d'un axe contenu dans le plan XOY

- $U_0 = \frac{U}{Z_0}$, $V_0 = \frac{V}{Z_0}$, $W_0 = \frac{W}{Z_0}$

	a	b	div	rot	hyp1	hyp2
T_{XY}	U_0	V_0	$-\gamma_1.U_0 - \gamma_2.V_0$	$-\gamma_1.V_0 + \gamma_2.U_0$	$-\gamma_1.U_0 + \gamma_2.V_0$	$-\gamma_1.V_0 - \gamma_2.U_0$
$T_{XY} + T_Z$	U_0	V_0	$-\gamma_1.U_0 - \gamma_2.V_0$	$-\gamma_1.V_0 + \gamma_2.U_0$	$-\gamma_1.U_0 + \gamma_2.V_0$	$-\gamma_1.V_0 - \gamma_2.U_0$
$T_{XY} + R_Z$	U_0	V_0	$-\gamma_1.U_0 - \gamma_2.V_0$	$2.C - \gamma_1.V_0 + \gamma_2.U_0$	$-\gamma_1.U_0 + \gamma_2.V_0$	$-\gamma_1.V_0 - \gamma_2.U_0$
$T_{XY} + R_{XY}$	$U_0 + B$	$V_0 - A$	$-\gamma_1.U_0 - \gamma_2.V_0$	$-\gamma_1.V_0 + \gamma_2.U_0$	$-\gamma_1.U_0 + \gamma_2.V_0$	$-\gamma_1.V_0 - \gamma_2.U_0$
T_Z	0	0	$-2.W_0$	0	0	0
$T_Z + R_{XY}$	B	$-A$	$-2.W_0$	0	0	0
R_Z	0	0	0	$2.C$	0	0
$R_Z + R_{XY}$	B	$-A$	0	$2.C$	0	0
$T_Z + R_Z$	0	0	$-2.W_0$	$2.C$	0	0
R_{XY}	$-A$	B	0	0	0	0

Tableau 1 : liens entre les différents mouvements 3D et les descripteurs

Dans ce tableau, on constate que 5 classes principales ressortent, ce qui permet de définir un jeu de cinq étiquettes décrivant de façon assez complète les différents mouvements les plus caractéristiques d'un solide, dans la base $(div, rot, hyp1, hyp2)$:

- $\Theta_1 = (0, 0, 0, 0)$: translation parallèle au plan image d'une surface parallèle à ce plan, ou rotation par rapport à l'un des axes X ou Y.
- $\Theta_2 = (div, 0, 0, 0)$: translation le long de l'axe de visée.
- $\Theta_3 = (0, rot, 0, 0)$: rotation autour de l'axe de visée.
- $\Theta_4 = (div, rot, 0, 0)$: composition d'une translation suivant l'axe de visée et d'une rotation autour de cet axe.
- $\Theta_5 = (div, rot, hyp1, hyp2)$: autre mouvement.

On peut illustrer la représentativité de ce jeu d'étiquettes en se plaçant par exemple dans un contexte routier. Un véhicule animé d'un mouvement frontal par rapport au capteur aura pour étiquette $\Theta_2(div, 0, 0, 0)$. Si la caméra est fixe, un mouvement transverse sera étiqueté par $\Theta_1(0, 0, 0, 0)$, de même qu'un mouvement de rotation autour d'un axe vertical. L'étiquette $\Theta_5(div, rot, hyp1, hyp2)$ caractérisera par exemple des véhicules animés d'un mouvement de translation non axiale et/ou non transverse, composé d'un mouvement de rotation.

3 Processus d'étiquetage

On se place dans une région de l'image où un seul type de mouvement est supposé présent. Il est donc nécessaire de réaliser une segmentation spatio-temporelle préalable de l'image. Cette étape indispensable a pour but de partitionner l'image en régions cohérentes pour le mouvement, chacune d'elle étant représentée par un modèle de vitesse linéaire unique. La méthode est décrite dans [1].

Disposant de cinq classes, donc de cinq modèles, le problème d'étiquetage à résoudre est en fait un problème de validation de modèles. L'approche adoptée pour réaliser cette validation est une approche statistique basée sur des critères d'information.

3.1 Présentation des critères d'information utilisés

Trois critères ont été utilisés, les critères d'Akaike, de Takeuchi et de Rissanen. Ces critères sont dérivés d'approches différentes. Les critères d'Akaike

et de Takeuchi sont obtenus à la suite d'un développement de la distance de Kullback-Leibler entre la densité de probabilité vraie inconnue f_0 d'une variable aléatoire multidimensionnelle X_n , et un modèle $f(\cdot; \Theta)$, Θ représentant le vecteur de paramètres de cette distribution. Le détail du développement est présenté en Annexe 1. Le critère de Rissanen est en fait déduit dans un autre contexte, [10]. Il s'agit dans ce cas d'un problème initial de codage où l'objectif est d'atteindre une description binaire des données de longueur minimale. Il a été par ailleurs montré qu'une résolution bayésienne de ce problème de sélection de modèles aboutit aussi à ce critère en conditions asymptotiques [11] (Annexe 2).

Ces critères, fonction du vecteur de paramètres Θ , ont la forme suivante, [6, 12] :

$$-\log[f(x_n; \Theta)] + P(\Theta) \quad (8)$$

où $P(\Theta)$ est un terme de pénalisation sur la complexité du modèle. Le premier terme de ce critère est un terme de vraisemblance. Selon la façon dont on dérive le critère, le terme de pénalisation $P(\Theta)$ prend différentes expressions, et l'on aboutit donc aux différents critères :

- Akaike : $AIC = -2.\log[f(x_n; \Theta)] + 2.\dim(\Theta)$ (9)

- Takeuchi : $TIC = -2.\log[f(x_n; \Theta)] + 2.\text{trace}(I_n(\Theta) * J_n(\Theta)^{-1})$ (10)

- Rissanen : $RIC = -2.\log[f(x_n; \Theta)] + \dim(\Theta).\log(n)$ (11)

où $I_n(\Theta)$ désigne la matrice d'information de Fisher, $J_n(\Theta)$ l'opposée de la matrice du Hessian.

Ces critères traduisent en fait l'idée intuitive suivante : si l'amélioration apportée par un modèle donné n'est pas suffisamment significative vis-à-vis d'un modèle moins complet, on choisira ce dernier.

Dans le cadre de cette étude, la famille de modèles considérée est composée, comme défini plus haut, de cinq modèles : $(0, 0, 0, 0)$, $(div, 0, 0, 0)$, $(0, rot, 0, 0)$, $(div, rot, 0, 0)$ et $(div, rot, hyp1, hyp2)$, de dimensions respectives 2, 3, 3, 4 et 6. On cherche donc à sélectionner, parmi le lot d'étiquettes, soit parmi les modèles considérés, celui qui décrit le mieux le mouvement 3D de la surface correspondant à la région analysée.

3.2 Formulation explicite des critères

On considère la variable aléatoire scalaire $e_\Theta(x, y)$, ainsi définie :

$$e_\Theta(x, y) = (\vec{\omega}_\Theta - \vec{\omega}).\vec{\nabla} f = \vec{\omega}_\Theta.\vec{\nabla} f - \vec{\omega}.\vec{\nabla} f \quad (11)$$

où

$\vec{\omega}$ désigne le vecteur vitesse du point (x, y) ,

$\vec{\omega}_\Theta$ le vecteur vitesse issu du modèle au point (x, y) ,

$\vec{\nabla} f(x, y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ le gradient spatial de la fonction intensité.

Deux cas se présentent : soit l'on dispose d'une estimation des vecteurs vitesse réels, calculés préalablement, soit on évite cette estimation. Dans ce dernier cas, afin d'explicitier $e_\Theta(x, y)$, on recourt à l'équation issue de l'approche différentielle, reliant vecteur vitesse et gradients spatio-temporels de la fonction intensité [2], c'est-à-dire : $\vec{\omega} \cdot \vec{\nabla} f = -\frac{\partial f}{\partial t}$, où $\frac{\partial f}{\partial t}(x, y)$ représente le gradient temporel de la fonction intensité; on a alors :

$$e_\Theta = \vec{\omega}_\Theta \cdot \vec{\nabla} f + \frac{\partial f}{\partial t} \quad (12)$$

On suppose que les variables e_Θ en chaque point $p_i(x_i, y_i)$ ($i = 1, n$), de la région R , sont indépendantes entre elles, et distribuées suivant une même loi normale centrée, de variance σ^2 .

La vraisemblance de l'observation conditionnellement à un vecteur de paramètres Θ vaut alors :

$$f((e(p_1), \dots, e(p_n)); \Theta) = \prod_{(x_i, y_i) \in R} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{e_\Theta^2(x_i, y_i)}{2\sigma^2} \right) \quad (13)$$

On considère en fait la fonction suivante :

$$\begin{aligned} \Psi(\Theta) &= -\log(f_\Theta(e(p_1), \dots, e(p_n))) \\ &= \frac{n}{2} \log(2\pi) + \frac{n}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \sum_{(x_i, y_i) \in R} e_\Theta^2(x_i, y_i) \end{aligned} \quad (14)$$

que l'on désire minimiser par rapport à Θ . Soit $\hat{\Theta}$ le vecteur de paramètres optimaux. Considérant que la variance dépend du modèle, on en fait une estimation a posteriori :

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{(x_i, y_i) \in R} e_\Theta^2(x_i, y_i) \quad (15)$$

Par conséquent, en tenant compte de (15), on aboutit à l'expression suivante:

$$\Psi(\hat{\Theta}) = \frac{n}{2} \log(2\pi) + \frac{n}{2} \log(\widehat{\sigma^2}) + \frac{n}{2} \quad (16)$$

Pour chaque modèle donné m , on recherche tout d'abord le vecteur $\hat{\Theta}^m$ optimal (qui minimise (14)); puis on choisit le modèle qui minimise : $\Psi(\hat{\Theta}^m) + P(\hat{\Theta}^m)$. Plus précisément, retenir le modèle le plus représentatif au sens de l'un des critères choisis reviendra à retenir le modèle de vecteur de paramètres donnés par:

$$\hat{\Theta}_{AIC} = \arg \min_{\hat{\Theta}^m} \left[\log \left(\sum_{(x_i, y_i) \in R} e_{\hat{\Theta}^m}^2(x_i, y_i) \right) + 2 \frac{\dim(\hat{\Theta}^m)}{n} \right]$$

ou

$$\hat{\Theta}_{TIC} = \arg \min_{\hat{\Theta}^m} \left[\log \left(\sum_{(x_i, y_i) \in R} e_{\hat{\Theta}^m}^2(x_i, y_i) \right) + 2 \frac{\text{trace}(\hat{I}_n(\hat{\Theta}^m) * \hat{J}_n(\hat{\Theta}^m)^{-1})}{n} \right]$$

ou

$$\hat{\Theta}_{RIC} = \arg \min_{\hat{\Theta}^m} \left[\log \left(\sum_{(x_i, y_i) \in R} e_{\hat{\Theta}^m}^2(x_i, y_i) \right) + \log(n) \frac{\dim(\hat{\Theta}^m)}{n} \right]$$

Les matrices $I_n(\hat{\Theta}^m)$ et $J_n(\hat{\Theta}^m)$ sont estimées ainsi :
on note

$$l_i(\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{e_{\Theta}^2(i)}{2\sigma^2}$$

on a alors, pour p et $q \in [1, \dim(\Theta^m)]$

$$\hat{I}_n(p, q) = \sum_{i=1 \rightarrow n} \frac{\partial l_i}{\partial \Theta_p^m} * \frac{\partial l_i(\hat{\Theta}^m)}{\partial \Theta_q^m}(\hat{\Theta}^m)$$

$$\hat{J}_n(p, q) = - \sum_{i=1 \rightarrow n} \frac{\partial^2 l_i}{\partial \Theta_p^m \partial \Theta_q^m}(\hat{\Theta}^m)$$

3.3 Evaluation de $\hat{\Theta}^m$

Dans le cas où l'on ne dispose pas d'une mesure du champ de vitesse réel, on considère donc l'observation e_{Θ^m} , dont l'expression explicite est la suivante :

$$\begin{aligned} e_{\Theta}(x, y) = & a_g \cdot f_x + b_g \cdot f_y + f_t + \\ & \frac{1}{2} \cdot \text{div}. [(x - x_g) \cdot f_x + (y - y_g) \cdot f_y] + \\ & \frac{1}{2} \cdot \text{rot}. [(x - x_g) \cdot f_y - (y - y_g) \cdot f_x] + \\ & \frac{1}{2} \cdot \text{hyp1}. [(x - x_g) \cdot f_x - (y - y_g) \cdot f_y] + \\ & \frac{1}{2} \cdot \text{hyp2}. [(x - x_g) \cdot f_y + (y - y_g) \cdot f_x] \end{aligned}$$

où $f_x = \frac{\partial f}{\partial x}$, $f_y = \frac{\partial f}{\partial y}$, $f_t = \frac{\partial f}{\partial t}$.

Dans le cas par exemple du modèle simple Θ_1 , cette observation se réduit à :

$$e_{\Theta}(x, y) = a_g \cdot f_x + b_g \cdot f_y + f_t$$

Maximiser la vraisemblance revient à minimiser la fonction $\Psi(\Theta)$, ce qui revient à rechercher le vecteur Θ minimisant

$$\sum_{(x_i, y_i) \in R} e_{\Theta}^2(x_i, y_i)$$

Ainsi, pour le modèle le plus complet Θ_5 , il faudra résoudre le système :

$$\left\{ \begin{array}{l} \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial a_g} = 0 \\ \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial b_g} = 0 \\ \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial div} = 0 \\ \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial rot} = 0 \\ \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial hyp1} = 0 \\ \sum_R e_{\Theta}(x, y) \cdot \frac{\partial e_{\Theta}(x, y)}{\partial hyp2} = 0 \end{array} \right.$$

ce qui peut s'écrire sous la forme matricielle :

$$M \cdot \Theta = R$$

Pour les modèles Θ_1 à Θ_5 , M sera respectivement de dimension $2*2$, $3*3$, $3*3$, $4*4$, $6*6$.

Une démarche analogue est adoptée si l'on dispose des champs de vitesse, avec cette fois-ci : $e_{\Theta}(x, y) = (\vec{\omega}_{\Theta} - \vec{\omega}) \cdot \vec{\nabla} f$.

4 Résultats

Bien qu'ayant testé les trois critères évoqués précédemment, les résultats présentés ci-dessous ne concernent que les critères d'Akaike et de Rissanen. D'une part, le critère de Takeuchi est plus lourd à mettre en place. D'autre

part, il est apparu qu'il ne fournissait pas des résultats concluant pour le problème présent. En fait, il semble confirmer le rôle du terme de vraisemblance, c'est-à-dire que dans tous les cas, il pénalisera non pas les modèles les plus complets, mais les modèles qui conviennent le moins aux données. Or il est évident que c'est le modèle complet qui est favorisé par un tel comportement.

La méthode a été testée à la fois sur des données synthétiques, et sur des données réelles avec mouvements synthétiques. Les résultats obtenus sur des champs de vecteurs simulés perturbés par un bruit additif sont parfaits, aussi bien avec le critère d'Akaike que de Rissanen : l'étiquette choisie est toujours celle qui correspond au mouvement simulé, même avec des rapports Signal/Bruit très faibles (figures 1a et 1b). Sur des champs de gradients simulés, on obtient de moins bons résultats. Souvent, avec les deux critères, c'est le modèle le plus complet (*div, rot, hyp1, hyp2*) qui est validé, même lorsque le vrai modèle est plus simple. Cependant, la valeur du critère pour la bonne étiquette ne présente jamais un écart important avec celle de l'étiquette complète, alors que les étiquettes ne convenant absolument pas (par exemple $(0, rot, 0, 0)$ dans le cas d'un mouvement frontal pur par rapport au capteur) présentent des valeurs très éloignées (figures 2, 3, 4).

On a aussi testé la méthode sur des images réelles quant à leur contenu. Ces images sont formées de quatre versions d'une fenêtre extraite d'une image réelle, et soumise à des déplacements synthétiques : dilatation, translation, rotation, composition d'une rotation et d'une dilatation. Ces images sont présentées sur les figures (5) et (6). L'étiquetage a été effectué sur les régions issues de la phase de segmentation spatio-temporelle, effectuée sur l'image composée (figure (7)). Les résultats sont présentés sur la figure (8). Le critère de Rissanen donne de meilleurs résultats que celui d'Akaike. En effet, ce dernier valide systématiquement le modèle complet, quelque soit la région analysée. Par le critère de Rissanen, on parvient à étiqueter correctement trois régions ; pour la région 3, dont le mouvement est en réalité un mouvement de rotation dans le plan image, donc de modèle $(0, rot, 0, 0)$, c'est le modèle (*div, rot, 0, 0*) qui a été validé, le vrai modèle présentant la seconde plus petite valeur de critère.

D'une façon générale, les paramètres sont sous-estimés. Néanmoins, on constate une amélioration de l'estimation lorsqu'on effectue un filtrage gaussien des images.

Si l'on ne regarde que le terme de vraisemblance, on constate qu'il permet de décider de la présence ou l'absence de tel ou tel descripteur dans le champ de vitesses d'une région. Dans le cas par exemple où la région analysée correspond à un objet subissant un mouvement de rotation autour de l'axe de visée, composé ou non à un autre mouvement, les étiquettes ne contenant

pas le descripteur *rot* présentent les valeurs de critère les plus fortes : si l'on évalue la vraisemblance pour le modèle complet $(div, rot, hyp1, hyp2)$, et que l'on force ensuite le terme *rot* à 0, on constate que la vraisemblance du modèle $(div, 0, hyp1, hyp2)$ est bien plus faible que la vraisemblance du modèle le plus complet. Ceci permet donc de conclure que le terme de *rot* est bien présent dans le champ de vecteurs de la région.

Les commentaires que l'on peut tirer de l'ensemble de ces résultats sont les suivants : les critères utilisés sont intéressants car ils permettent de tester tous les modèles conjointement, sans se confronter à des problèmes de seuillage. Grâce au terme de vraisemblance, ils permettent de discriminer les étiquettes ne convenant pas du tout au champ de vitesses de la région analysée, de celles qui conviennent. Cependant, ils sont plutôt plats, en partie à cause du peu de différence de dimension entre les différents modèles. De plus, la vraie distribution n'appartient probablement pas à la famille de modèles (on ne propose donc qu'une approximation de cette distribution). Ceci s'ajoutant au fait que l'on se contente d'un développement du champ de vitesses au premier ordre, il apparaît que le terme de pénalisation ne joue pas un rôle assez significatif par rapport au terme de vraisemblance. D'autre part, il est important de travailler sur de bonnes estimées, ce qui justifie la nécessité d'une bonne phase de segmentation. Cette segmentation a pour effet de découper l'image en régions représentées par un unique modèle de vitesse. Elle permet aussi d'éliminer les zones aberrantes, ou pauvres en information de l'image, permettant ainsi une meilleure estimation des paramètres. En dernier lieu, l'estimation des gradients spatio-temporels est une phase très importante. Le préfiltrage des images améliore notablement la qualité de l'estimation des paramètres, mais comme il détruit en partie l'information contenue dans les images, il est préférable de réaliser la segmentation sur des données peu filtrées, et de filtrer les cartes de gradients spatio-temporels après cette phase de segmentation.

5 Conclusion

Cette étude est une première étape vers une interprétation qualitative du mouvement. La démarche adoptée a été de rechercher des outils simples, en l'occurrence des opérateurs différentiels comme la divergence, le rotationnel, bien adaptés pour qualifier des mouvements rigides 3D. Ces descripteurs ont permis de définir un jeu d'étiquettes symboliques pouvant décrire de façon assez complète les différents mouvements possibles. L'étape d'étiquetage a été abordée statistiquement, en se basant sur les critères d'information d'Akaike et de Rissanen. Ces descripteurs ont déjà été utilisés, notamment dans [15] où les termes *div*, *rot* et $(hyp1^2 + hyp2^2)$ sont comparés à 0. L'apport de cette étude se situe sur deux plans : d'une part, on a directe-

ment relié les descripteurs au mouvement apparent et au mouvement 3D, justifiant ainsi explicitement leur importance. D'autre part, on a utilisé une méthode de décision évitant la phase de seuillage peu aisée dans la mesure où il n'est pas évident de décider dans l'absolu jusqu'à quel niveau la dynamique d'un descripteur est significative.

Le critère de Rissanen a donné les résultats les plus encourageants. Cependant, ces résultats sont loin d'être parfaits. Trois raisons principales semblent pouvoir l'expliquer :

- on ne travaille pas directement sur un champ de vecteurs, mais sur des variables scalaires e_Θ , qui dépendent des gradients spatio-temporels. L'estimation de ces derniers est imparfaite, perturbant notablement l'estimation de paramètres.
- les critères sont relativement plats, et ne discriminent donc pas fortement les modèles adéquats des autres.
- la dernière remarque comporte deux points. Premièrement, on commet une erreur d'estimation des paramètres, qui provoque une déformation du champ mesuré vis-à-vis du champ de vitesse véritable. Cette déformation est sûrement différente d'un simple bruit additif gaussien centré. Deuxièmement, on se trouve dans une situation où les distances inter-modèles sont quasi-nulles entre les mouvements "purs" et le mouvement complexe (étiquette Θ_5). D'où une tendance à valider le modèle le plus complet dès qu'on s'écarte d'une estimation parfaite. Pour y remédier, il faudrait tenir compte de l'incertitude sur $\Psi(\hat{\Theta})$ ou introduire une notion "élargie" de mouvement quasi-translational, quasi-divergent, etc...

Un autre problème crucial auquel on est confronté est celui de la segmentation spatio-temporelle. La méthode utilisée est une méthode de division-fusion, qui présente le défaut d'être peu stable : de petites variations de seuil provoquent des changements notables dans la partition de l'image. De même, pour deux images successives, la segmentation obtenue peut différer sensiblement. Ceci peut aussi s'expliquer par la qualité et la complexité des séquences traitées.

Les développements en cours s'orientent vers quatre axes :

- dans le cadre du projet européen PROMETHEUS, concernant l'aide à la conduite automobile, on va désormais traiter des séquences réelles en contexte routier.
- pour pallier le problème de la segmentation spatio-temporelle, une approche différente peut être envisagée : au lieu de procéder à cette

segmentation, on peut directement aborder le problème d'étiquetage, par une approche bayésienne et via une modélisation markovienne, en tenant compte d'informations contextuelles et temporelles a priori. L'étiquetage serait géré localement. Afin d'éviter d'avoir à traiter un nombre de données trop important (en l'occurrence l'image entière), on envisage de procéder à une étape préalable de détection des zones en mouvement apparent, [5]. Un tel prétraitement doit de plus permettre d'éliminer les zones de l'image ne contenant pas d'information.

- on peut aussi envisager une approche arborescente. Dans un problème de détection d'obstacles dans un contexte routier par exemple, donc avec un capteur mobile dont l'axe de visée correspond à l'axe de déplacement, le terme de divergence est alors prépondérant. Il est dans ce cas envisageable de construire un arbre de décision s'articulant autour de ce terme *div*.
- en dernier lieu, il est évident que l'aspect temporel doit être intégré dans ce problème d'analyse de scène. Ceci peut à la fois améliorer l'étiquetage, et le stabiliser au cours du temps. On peut d'ailleurs intégrer cet aspect temporel dans le modèle markovien évoqué ci-dessus.

Bibliographie

- [1] P. BOUTHEMY, J. SANTILLANA RIVERO : "Segmentation en régions selon des critères de mouvement dans une séquence d'images", Proc. 6ème Congrès AFCET RF et IA, pp.105-114, 1987.
- [2] B.K.P. HORN, B.G. SCHUNCK : "Determining Optical Flow", Artificial Intelligence, Vol. 17, 1981, pp. 185-203.
- [3] J.J. KOENDERINK : "Optic flow", Vision Res., Vol. 36, No. 1, 1986, pp.161-180.
- [4] J.J. KOENDERINK, A.J. VAN DOORN : "Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer", Optica Acta, Vol. 22, No. 9, 1975, pp 773-791.
- [5] P. LALANDE, P. BOUTHEMY : "Détection de zones en mouvement dans une séquence d'images selon une approche markovienne", GRETSI, pp 707-710, 12-16 Juin 1989.
- [6] L. LJUNG : "System identification theory for the user", Prentice-Hall, Information and System Sciences series, 1987, Chap.7.
- [7] H.C. LONGUET-HIGGINS, K. PRAZDNY : "The interpretation of a moving retinal image", Proc. Roy. Soc. Lond., Vol. B-208, 1980, pp 385-397.
- [8] H.H. NAGEL : "From image sequences towards conceptual descriptions", Image and vision computing, Vol. 6, No. 2, Mai 1988, pp 59-74. pp 1148-1151.
- [9] B. NEUMANN, H.J. NOVAK : "Event models for recognition and natural language description of events in real-world image sequences", Proc. IJCAI, 1983, pp 724-726.
- [10] J. RISSANEN : "Modeling by shortest data description". Automatica, vol.14, 1986, pp.465-472.
- [11] G. SCHWARZ : "Estimating the dimension of a model". Ann. Statist., vol.6, 1978, pp.461-464.
- [12] R. SHIBATA : "Criteria of statistical model selection", Research Report, KSTS/RR-86/009, Août 1986, Dept. of Math., Keio Univ., Yokohama, Japon.

- [13] P.Y. SIMARD, G.E. MAILLOUX : "A projection operator for the restoration of divergence-free vector-fields", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-10, No.2, Mars 1988, pp 248-256.
- [14] M. SUBBARAO : "Bounds on translational and angular velocity components from first order derivatives of image flow", Proc. AAAI, 1987, pp 744-748.
- [15] W.B. THOMPSON, K.M. MUTCH, V.A. BERZINS : "Analyzing object motion based on optical flow", Proc. ICPR, 1984, pp.791-794.
- [16] J.K. TSOTSOS, J. MYLOPOULOS, H.D. COVVEY, S.W.ZUCKER : "A framework for visual motion understanding", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 6, Nov. 1980, pp 563-573.
- [17] J. ZHANG, J.W. MODESTINO : "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation", Proc. ICASSP, 1988, pp 1148-1151.

ETIQUETAGE SUR CHAMP DE VITESSE SIMULE

Etiquetage sur champ de vitesse perturbé par un bruit blanc additif. Le champ est un champ divergent pur (modèle Θ_2).

Légende :

- ◇ critère(Θ_1)-critère(Θ_5)
- + critère(Θ_2)-critère(Θ_5)
- critère(Θ_3)-critère(Θ_5)
- × critère(Θ_4)-critère(Θ_5).

L'étiquetage a été effectué sur une fenêtre 25*25 non centrée sur le foyer d'expansion.

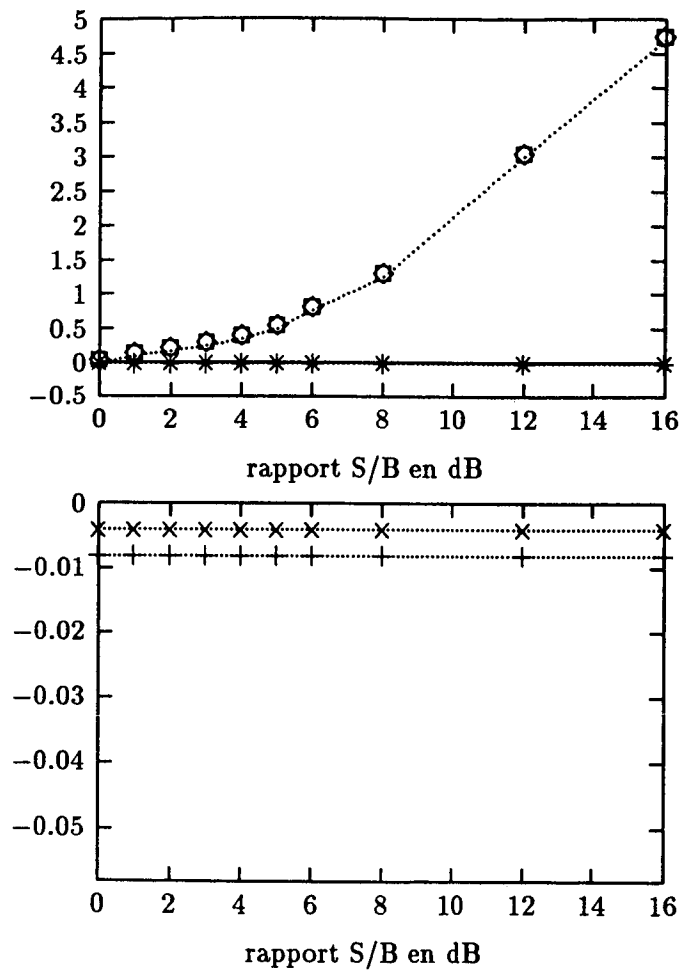


Figure 1a : Critère d'Akaike

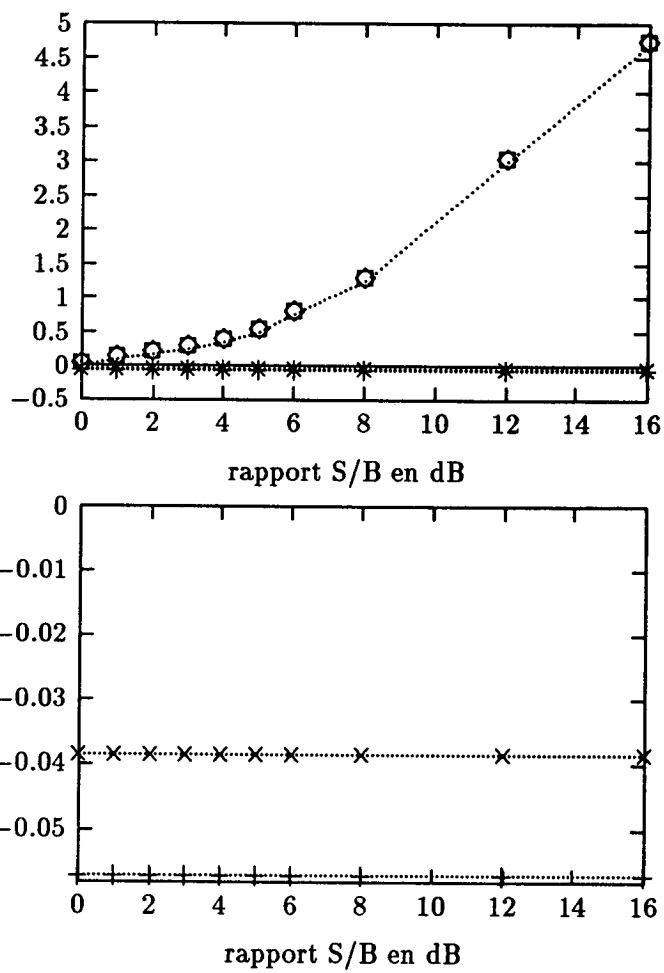


Figure 1b : Critère de Rissanen

ETIQUETAGE SUR CHAMPS DE GRADIENTS SIMULES

Etiquetage sur gradients : on a généré des champs de gradients spatio-temporels parfaits, c'est-à-dire vérifiant parfaitement l'équation différentielle liant l'intensité lumineuse au vecteur vitesse, pour un mouvement divergent pur (modèle Θ_2). Différents types de bruit y ont été ajoutés.

Légende :

- ◇ critère(Θ_1)-critère(Θ_5)
- + critère(Θ_2)-critère(Θ_5)
- critère(Θ_3)-critère(Θ_5)
- × critère(Θ_4)-critère(Θ_5).

L'étiquetage a été effectué sur une fenêtre 25*25 non centrée sur le foyer d'expansion.

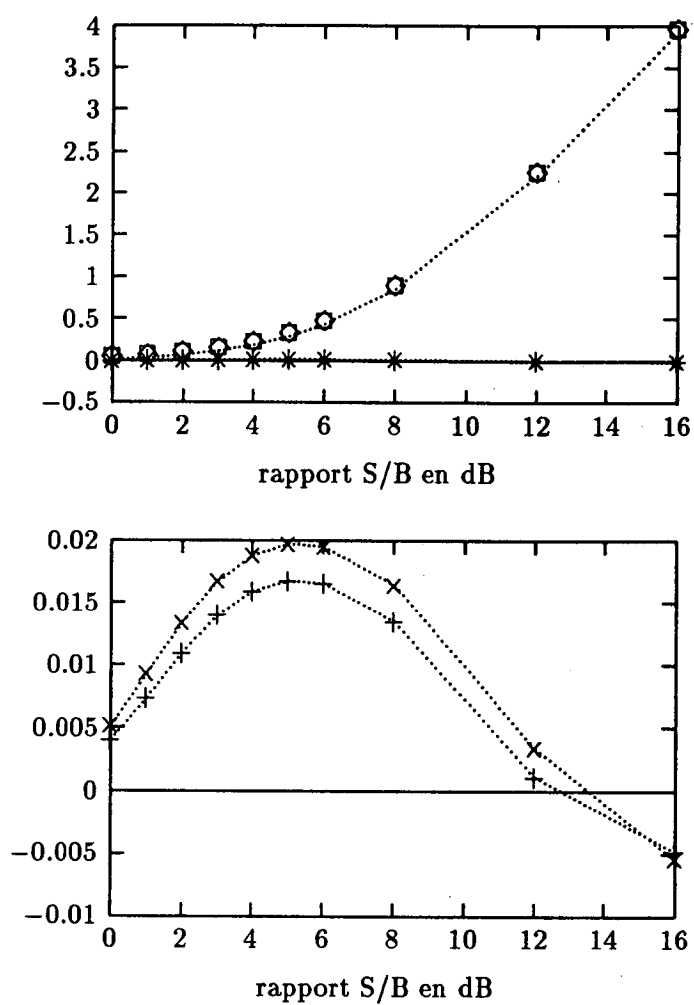


Figure 2a : Bruit additif non corrél  entre les gradients spatiaux et temporel. Crit re d'Akaike

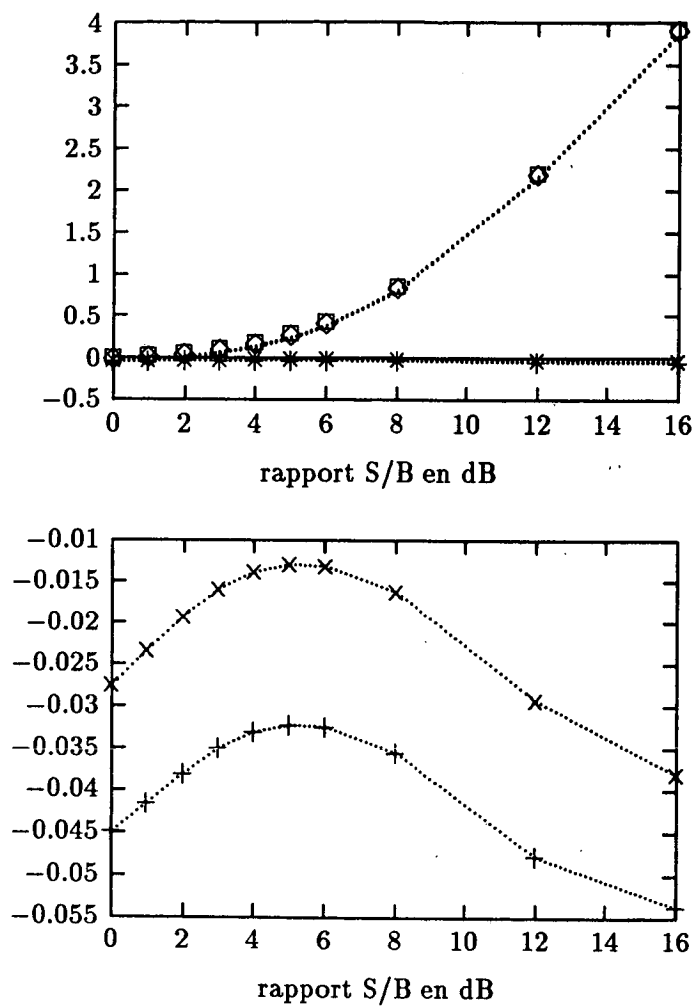


Figure 2b : Bruit additif non corrél  entre les gradients spatiaux et temporel. Crit re de Rissanen

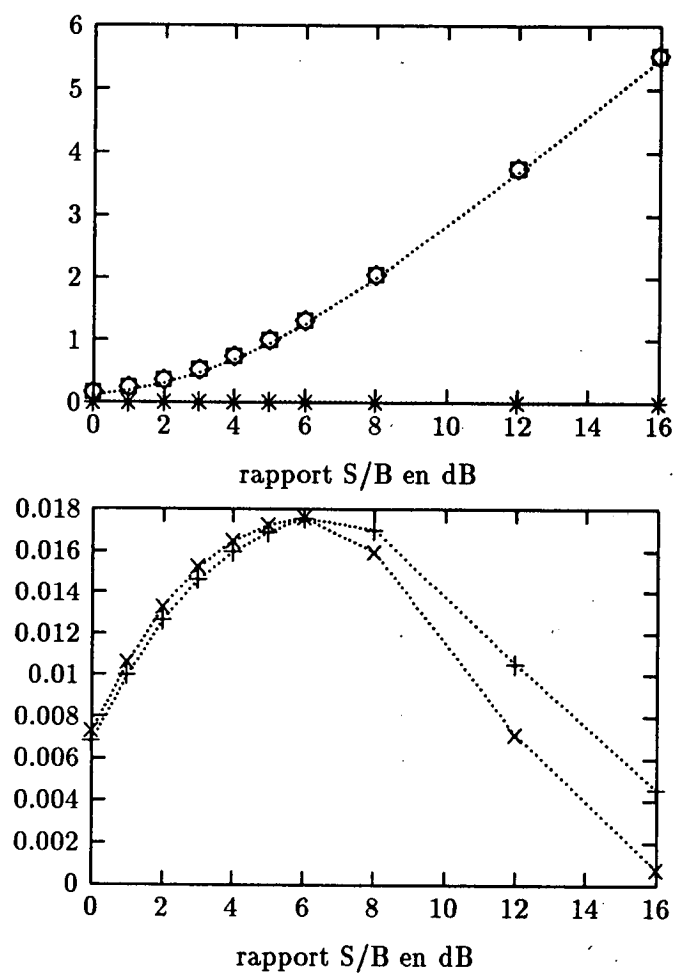


Figure 3a : Bruit additif corrélé entre les gradients spatiaux et temporel.
Critère d'Akaike

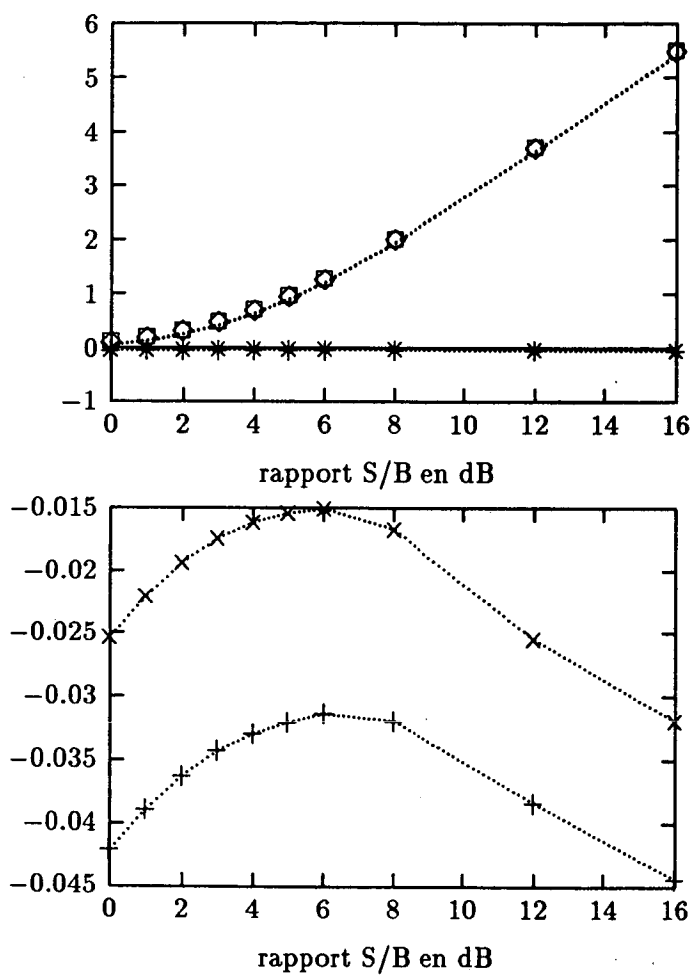


Figure 3b : Bruit additif corrélé entre les gradients spatiaux et temporel.
Critère de Rissanen

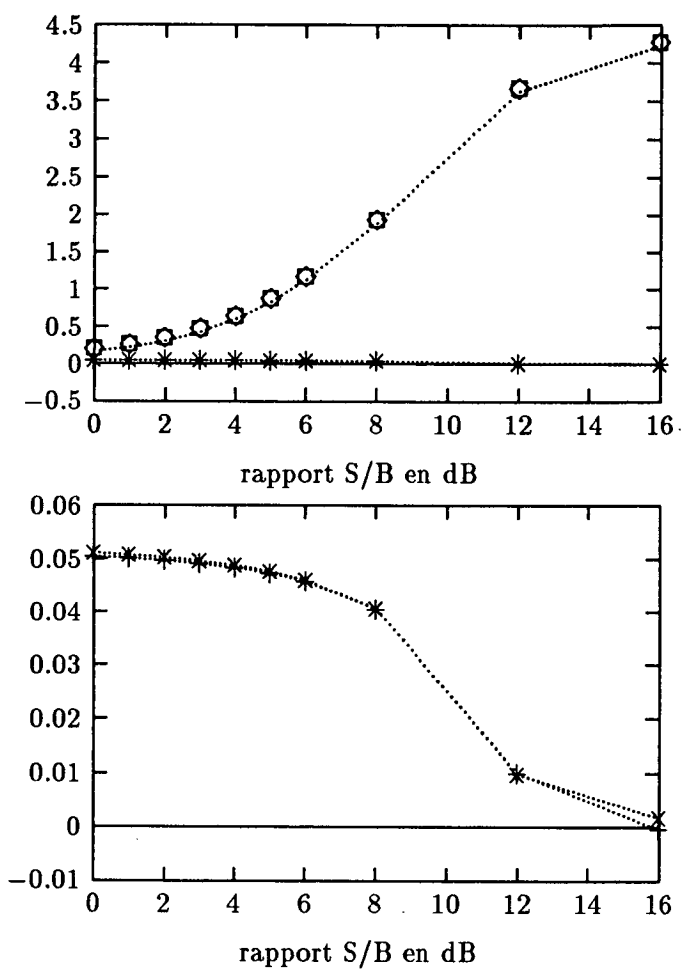


Figure 4a : Bruit additif non corrélé entre les gradients spatiaux et temporel avec 20 % de points aberrants. Critère d'Akaike

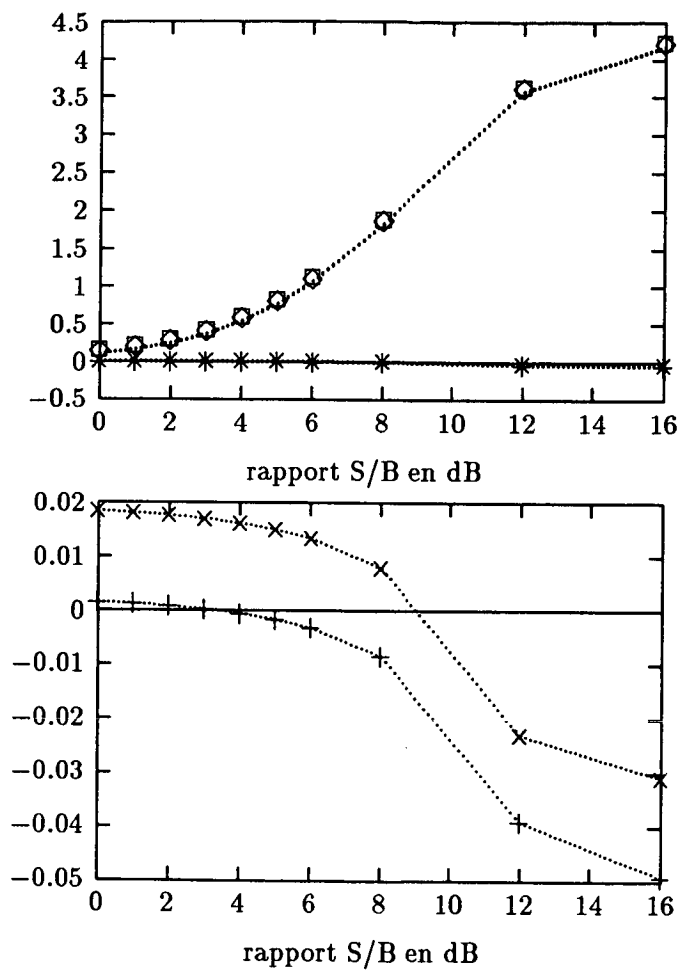


Figure 4b : Bruit additif non corrél  entre les gradients spatiaux et temporel avec 20 % de points aberrants. Crit re de Rissanen

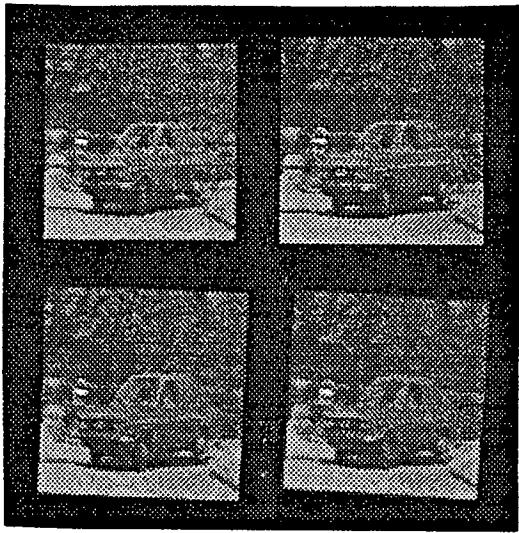


figure 5 : 1ère image

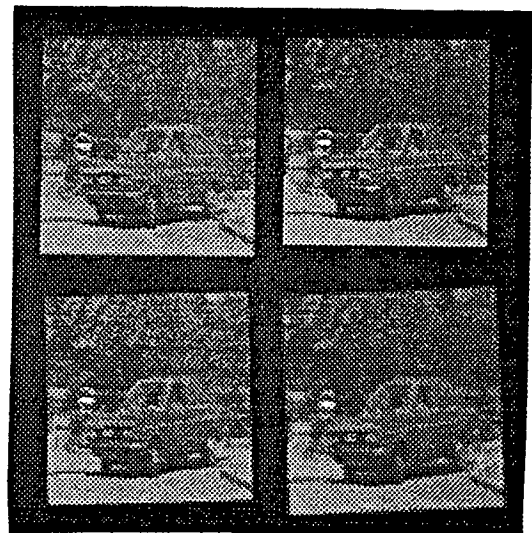


figure 6 : 2ème image

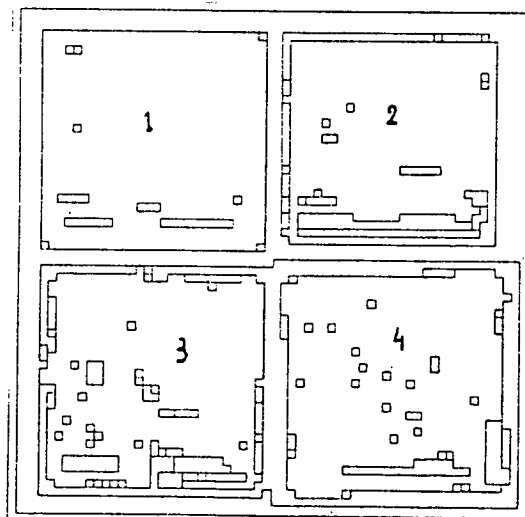


figure 7 : résultat de la segmentation spatio-temporelle

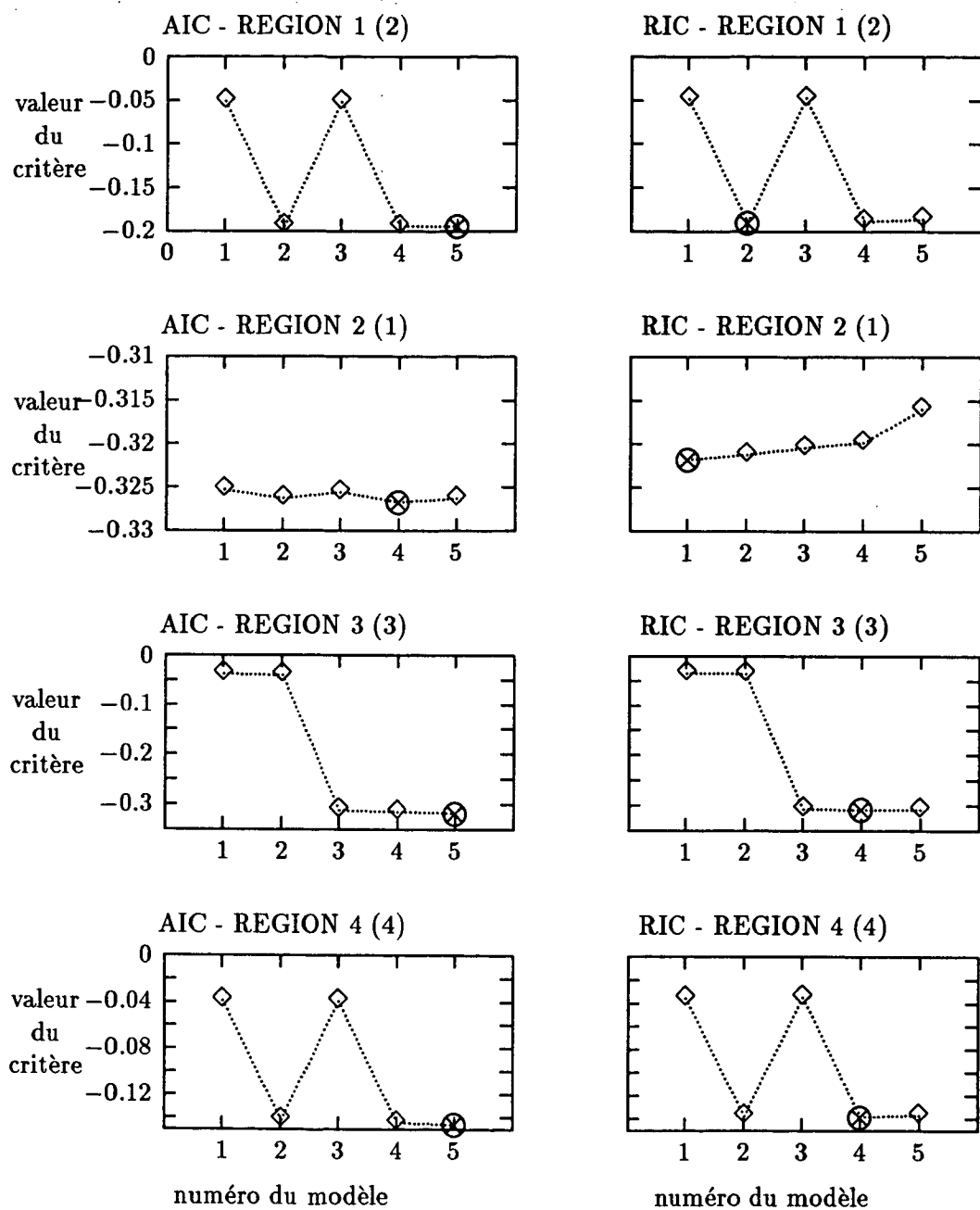


Figure 8 : Valeurs des critères pour les différents modèles, pour les 4 régions analysées. \otimes signale le modèle validé, \diamond les autres modèles. Les numéros de modèles sont ceux qui ont été donnés lors de la définition des étiquettes. Pour chaque région, le vrai modèle est indiqué entre parenthèses.

1 Annexe 1 : Critères d'Akaike et de Takeuchi, [6]

Soit X_n une variable aléatoire multidimensionnelle de dimension n , de densité inconnue $f_0(\cdot)$, et x_n une réalisation observable de cette variable. L'objectif est de rechercher parmi une famille de modèles paramétrés $F = [f(\cdot; \Theta_p)]$, Θ_p désignant le vecteur de paramètres, de dimension p , la densité la plus proche de $f_0(\cdot)$. On va donc en fait rechercher la dimension p optimale, c'est-à-dire telle que $f(\cdot; \Theta_p)$ coïncide le mieux avec $f_0(\cdot)$. Une mesure de la distance entre les deux densités $f_0(\cdot)$ et $f(\cdot; \Theta_p)$ peut être définie par la distance de Kullback-Leibler :

$$I(f_0(\cdot), f(\cdot; \Theta_p)) = \int f_0(z_n) \cdot \log \frac{f_0(z_n)}{f(z_n; \Theta_p)} \cdot dz_n$$

soit encore

$$I(f_0(\cdot), f(\cdot; \Theta_p)) = \int f_0(z_n) \cdot \log(f_0(z_n)) \cdot dz_n - \int f_0(z_n) \cdot \log(f(z_n; \Theta_p)) \cdot dz_n \quad (A1 - 1)$$

On note

$$l(\Theta_p) = \log f(x_n; \Theta_p)$$

$\hat{\Theta}_p$ le vecteur optimal au sens du Maximum de vraisemblance, c'est-à-dire maximisant $l(\Theta_p)$.

$I(f_0(\cdot), f(\cdot; \hat{\Theta}_p))$ dépend de $\hat{\Theta}_p$, donc des réalisations x_n . On considère donc l'espérance suivant X_n de $I(f_0(\cdot), f(\cdot; \hat{\Theta}_p))$ qui peut s'écrire ainsi :

$$\int f_0(z_n) \cdot \log(f_0(z_n)) \cdot dz_n - E_{X_n} \left[\int f_0(z_n) \cdot \log(f(z_n; \hat{\Theta}_p)) \cdot dz_n \right]$$

On considère que le meilleur modèle $f(\cdot; \Theta_p)$ est celui qui minimise cette quantité. Ceci revient à maximiser par rapport à p

$$E_{X_n} \left[\int f_0(z_n) \cdot \log(f(z_n; \hat{\Theta}_p)) \cdot dz_n \right] \quad (A1 - 2)$$

Le problème qui se pose est donc d'estimer ce terme.

Soit Θ_p^* le vrai vecteur de paramètres optimaux, c'est-à-dire l'unique solution du système :

$$E_{X_n} \left[\frac{\partial l(\Theta_p)}{\partial \Theta_p^i} \right] = 0 \quad i = 1, \dots, p \quad (A1-3)$$

Θ_p^* est inconnu, et on ne dispose que d'une estimation $\hat{\Theta}_p$ de ce vecteur.

En développant le terme $l(\hat{\Theta}_p)$ autour de Θ_p^* , on obtient :

$$\begin{aligned} \log f(z_n; \hat{\Theta}_p) &= \log f(z_n; \Theta_p^*) + (\hat{\Theta}_p - \Theta_p^*)^T \left[\frac{\partial \log f(z_n; \Theta_p)}{\partial \Theta_p} \right]_{\Theta_p^*} + \\ &\quad \frac{1}{2} (\hat{\Theta}_p - \Theta_p^*)^T \left[\frac{\partial^2 \log f(z_n; \Theta_p)}{\partial \Theta_p^2} \right]_{\Theta_p^{**}} (\hat{\Theta}_p - \Theta_p^*) \end{aligned}$$

où Θ_p^{**} est un vecteur compris entre $\hat{\Theta}_p$ et Θ_p^* .

Par conséquent, en considérant l'espérance suivant X_n de cette expression, et en tenant compte de (A1-3), on arrive à :

$$\begin{aligned} \int f_0(z_n) \cdot \log f(z_n; \hat{\Theta}_p) \cdot dz_n &= \int f_0(z_n) \cdot \log f(z_n; \Theta_p^*) \cdot dz_n + \\ \frac{1}{2} (\hat{\Theta}_p - \Theta_p^*)^T &\left(\int f_0(z_n) \cdot \left[\frac{\partial^2 \log f(z_n; \Theta_p)}{\partial \Theta_p^2} \right]_{\Theta_p^{**}} \cdot dz_n \right) (\hat{\Theta}_p - \Theta_p^*) \end{aligned} \quad (A1-4)$$

Le terme

$$\int f_0(z_n) \cdot \left[\frac{\partial^2 \log f(z_n; \Theta_p)}{\partial \Theta_p^2} \right]_{\Theta_p^{**}} \cdot dz_n$$

est en fait l'opposée de la matrice de Hessian en Θ_p^{**} , notée $J_n(\Theta_p^{**})$. On propose une approximation de l'espérance suivant X_n de ce terme par l'expression suivante

$$\text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right)$$

où $I_n(\Theta_p^*)$ désigne la matrice d'information de Fisher, et s'exprime ainsi :

$$I_n(\Theta_p^*) = E_{X_n} \left[\frac{\partial l(\Theta_p^*)}{\partial \Theta_p^i} \frac{\partial l(\Theta_p^*)}{\partial \Theta_p^j} \right]$$

On aboutit donc, en considérant l'espérance suivant X_n de l'expression (A1-4), et en tenant compte de cette approximation, à l'expression suivante :

$$\begin{aligned} E_{X_n} \left[\int f_0(z_n) \cdot \log f(z_n; \hat{\Theta}_p) \cdot dz_n \right] &= \int f_0(z_n) \cdot \log f(z_n; \Theta_p^*) \cdot dz_n \\ &\quad - \frac{1}{2} \text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right) = \\ E_{X_n} \left[l(\Theta_p^*) \right] &- \frac{1}{2} \text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right) \quad (\text{A1-5}) \end{aligned}$$

D'autre part, on développe $l(\Theta_p^*)$ autour de $\hat{\Theta}_p$, en tenant compte du fait que

$$\left[\frac{\partial l(\Theta_p)}{\partial \Theta_p^i} \right]_{\hat{\Theta}_p} = 0 \quad i = 1, \dots, p$$

et on obtient :

$$E_{X_n} \left[l(\Theta_p^*) \right] = E_{X_n} \left[l(\hat{\Theta}_p) \right] - \frac{1}{2} \text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right) \quad (\text{A1-6})$$

Par conséquent :

$$\begin{aligned} E_{X_n} \left[\int f_0(z_n) \cdot \log f(z_n; \hat{\Theta}_p) \cdot dz_n \right] &= \\ E_{X_n} \left[l(\hat{\Theta}_p) \right] &- \text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right) \quad (\text{A1-7}) \end{aligned}$$

On peut donc proposer de sélectionner le modèle p maximisant

$$-l(\hat{\Theta}_p) + t_n(\widehat{\Theta}_p^*)$$

où $t_n(\widehat{\Theta}_p^*)$ est une estimée de $\text{trace} \left(I_n(\Theta_p^*) J_n(\Theta_p^*)^{-1} \right)$. Partant de l'hypothèse que X_n est un vecteurs d'observations indépendantes, on peut écrire

$$l(\Theta) = \sum_{q=1 \rightarrow n} l_q(\Theta)$$

Deux façons d'estimer le terme $t_n(\widehat{\Theta}_p^*)$, à partir d'hypothèse différentes, aboutissent aux critères d'Akaike et de Takeuchi : le critère d'Akaike s'obtient en considérant que l'une des densités de la famille F est la vraie densité $f_0(\cdot)$. Dans ce cas, $I_n(\Theta_p^*)$ et $J_n(\Theta_p^*)$ sont égales, et $t_n(\widehat{\Theta}_p^*) = p$. Si les modèles proposés ne sont qu'une approximation de la vraie densité $f_0(\cdot)$, Takeuchi propose une approximation du terme $t_n(\widehat{\Theta}_p^*)$.

On aboutit donc aux deux critères suivants :

- Akaike : $AIC = -2.l(\hat{\Theta}_p) + 2.p$
- Takeuchi : $TIC = -2.l(\hat{\Theta}_p) + 2.\text{trace}(\hat{I} * \hat{J}^{-1})$

avec

$$\hat{I} = \left[\sum_{q=1 \rightarrow n} \frac{\partial l_q(\hat{\Theta}_p)}{\partial \Theta_p^i} \sum_{q=1 \rightarrow n} \frac{\partial l_q(\hat{\Theta}_p)}{\partial \Theta_p^j} \right]$$

et

$$\hat{J} = \left[\sum_{q=1 \rightarrow n} \frac{\partial^2 l_q(\hat{\Theta}_p)}{\partial \Theta_p^i \partial \Theta_p^j} \right]$$

2 Annexe 2 : Critère d'information selon une approche bayésienne, [11]

On considère comme dans l'Annexe 1 une variable multidimensionnelle X_n de dimension n . On suppose de plus que les observations x_n^i ($i=1, \dots, n$) sont indépendantes entre elles et de même loi normale centrée $f(x; \Theta_p)$, et dépendent linéairement de Θ_p :

$$f(x; \Theta_p) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp - \frac{\left(Y_0(x) + \sum_{k=1 \rightarrow p} \Theta_p^k \cdot Y_k(x) \right)^2}{2\sigma^2}$$

C'est en particulier le cas de la variable e_Θ introduite dans le paragraphe 3.2.

On désire sélectionner, parmi tous les modèles Θ , le modèle Θ_p présentant la probabilité a posteriori la plus grande, c'est-à-dire maximisant :

$$P(\Theta \in R^p, x_n) = P(p) \cdot \int_{R^p} P_\Theta(x_n) \cdot d\Theta$$

où $P(p)$ est une connaissance a priori sur la complexité du modèle. Il est donc nécessaire d'évaluer le terme

$$P = \int_{R^p} P_\Theta(x_n) \cdot d\Theta$$

On considère que

$$\sigma^2 = \frac{1}{n} \sum_{i=1 \rightarrow n} \left(Y_0(x^i) + \sum_{k=1 \rightarrow p} \Theta_p^k \cdot Y_k(x^i) \right)^2 \quad (A2-1)$$

Ainsi,

$$P = e^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \int_{R^p} (\sigma^2)^{-\frac{n}{2}} \cdot d\Theta \quad (A2-2)$$

Il est possible d'écrire σ^2 sous la forme suivante :

$$\sigma^2 = \Theta^T \cdot M \cdot \Theta + 2 \cdot \Theta^T \cdot R + R_{00} \quad (A2-3)$$

où

$$M(r, s) = \frac{1}{n} \sum_{i=1 \rightarrow n} Y_r(x_n^i) \cdot Y_s(x_n^i) \quad (r, s = 1, \dots, p)$$

$$R(r) = \frac{1}{n} \sum_{i=1 \rightarrow n} Y_0(x_n^i) \cdot Y_r(x_n^i) \quad (r = 1, \dots, p)$$

$$R_{00} = \frac{1}{n} \sum_{i=1 \rightarrow n} (Y_0(x_n^i))^2$$

La matrice M est en fait la matrice de Hessian évoquée dans l'annexe A1, du modèle $f(\cdot; \hat{\Theta}_p)$, où $\hat{\Theta}_p$ est le vecteur optimal au sens du Maximum de Vraisemblance. Cette matrice est symétrique définie positive, donc diagonalisable. Soit D sa matrice diagonale, dont les valeurs propres valent λ_k ($k = 1, \dots, p$), Q la matrice de passage de M à D . De même, on définit les vecteurs $\eta = Q \cdot \Theta$, et $U = Q \cdot R$. On a donc

$$d\eta = d\Theta$$

$$\begin{aligned} \sigma^2 &= \Theta^T \cdot (Q^T \cdot D \cdot Q) \cdot \Theta + 2 \cdot \Theta^T \cdot R + R_{00} \\ &= \eta^T \cdot D \cdot \eta + 2 \cdot \eta^T \cdot U + R_{00} \\ &= \sum_{k=1 \rightarrow p} (\lambda_k \cdot \eta_k^2 + 2 \cdot u_k \cdot \eta_k) + R_{00} \\ &= \sum_{k=1 \rightarrow p} \lambda_k \cdot \left(\eta_k + \frac{u_k}{\lambda_k} \right)^2 + \left(R_{00} - \sum_{k=1 \rightarrow p} \frac{u_k^2}{\lambda_k} \right) \end{aligned} \quad (A2 - 4)$$

Le terme $\left(R_{00} - \sum_{k=1 \rightarrow p} \frac{u_k^2}{\lambda_k} \right)$ est en fait la variance minimale possible pour Θ , c'est-à-dire la variance $\hat{\sigma}^2$ pour le vecteur optimal $\hat{\Theta}$ au sens du Maximum de Vraisemblance. On note

$$\xi_k = \eta_k + \frac{u_k}{\lambda_k}, \quad d\xi_k = d\eta_k$$

$$\text{et } C = e^{-\frac{1}{2n}} (2\pi)^{-\frac{1}{2n}} e^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}}$$

Par conséquent

$$P = C \int_{R^p} \left(\sum_{k=1 \rightarrow p} \lambda_k \cdot \xi_k^2 + \hat{\sigma}^2 \right)^{-\frac{n}{2}} \cdot d\xi$$

On réalise alors le changement de variable

$$\chi_k = \sqrt{\frac{\lambda_k}{\hat{\sigma}^2}} \xi_k$$

$$P = \frac{C \hat{\sigma}^{\frac{p-n}{2}}}{\det(M)} \int_{R^p} \left(1 + \sum_{k=1 \rightarrow p} \chi_k^2 \right)^{-\frac{n}{2}} \cdot d\chi$$

Le terme $\sum_{k=1 \rightarrow p} \chi_k^2$ est en fait le carré de la norme du vecteur χ .

Théorème :

Si F une fonction continue de R^+ dans R , alors

$$\int_{R^+} F(\|\chi\|) . d\chi = S_p \int_{R^+} F(r) . r^{p-1} . dr$$

$$\text{où } S_p := \frac{2 \cdot \pi^{\frac{p}{2}}}{\Gamma(\frac{p}{2})}$$

$$\text{et } \Gamma(k) = \int_{R^+} e^{-x} . x^k . dx$$

D'autre part, il est possible de considérer que

$$\int_{R^+} (1+r^2)^{-\frac{n}{2}} . r^{p-1} . dr \sim \int_{R^+} e^{-\frac{n}{2} . r^2} . r^{p-1} . dr$$

Par le changement de variable $\rho = \frac{n}{2} . r^2$, on montre que ce dernier terme vaut

$$2^{\frac{p-2}{2}} n^{-\frac{p}{2}} \Gamma(\frac{p}{2})$$

On arrive donc pour P à l'expression suivante :

$$P = e^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \left[\frac{\hat{\sigma}^2 \frac{p-n}{2}}{\det(M)} n^{-\frac{p}{2}} (2\pi)^{\frac{p}{2}} \right]$$

Le modèle le plus représentatif sera le modèle minimisant $-\log [P(\Theta \in R^p, x_n)]$

donc minimisant

$$\begin{aligned} (n-p) \log \hat{\sigma}^2 + \log(\det(M)) + p \cdot \log n - p \cdot \log(2\pi) - \log(P(p)) = \\ -2 \cdot \log f(x_n; \widehat{\Theta}_p) + p \cdot \log n + (\log(\det(M)) - p \log \hat{\sigma}^2 - p \cdot \log(2\pi)) \end{aligned}$$

En conditions asymptotiques, c'est-à-dire lorsque $n \rightarrow \infty$, on retrouve le critère de Rissanen :

$$-2 \cdot \log f(x_n; \widehat{\Theta}_p) + p \cdot \log n$$

LISTE DES DERNIERES PUBLICATIONS INTERNES IRISA

- PI 476 **SIMULATION REPARTIE DE SYSTEMES A EVENEMENTS DISCRETS:
PARTIE 1 : MODELISATION ET SCHEMAS D'EXECUTION**
Philippe INGELS, Michel RAYNAL
26 Pages, Juin 1989.
- 477 **PROGRAMMING WITH MALI - UNIFICATION OR ORDERED TYPES**
Olivier RIDOUX
18 Pages, Juin 1989.
- PI 478 **CLASSIFICATION OF CONCEPTS DESCRIBED BY TAXONOMIC
PREORDONNANCE VARIABLES WITH MULTIPLE CHOICE**
Israël-César LERMAN, Philippe PETER
16 Pages, Juin 1989.
- PI 479 **A SIMPLE GRAPH CONSTRUCTION OF SEMILINEAR REACHABILITY
SETS OF VECTOR ADDITION SYSTEMS**
Gilles LESVENTES
16 Pages, Juin 1989.
- PI 480 **THE MODELLING SYSTEM PYRAMIDE AS AN INTERACTIVE
HELP FOR THE GUIDANCE OF THE INSPECTION VEHICLE
CENTAURE**
Philippe EVEN, Lionel MARCE
22 Pages, Juin 1989.
- PI 481 **VERS UNE INTERPRETATION QUALITATIVE DE COMPORTEMENTS
CINEMATQUES DANS LA SCENE A PARTIR DU MOUVEMENT
APPARENT**
Edouard FRANCOIS, Patrick BOUTHEMY
40 Pages, Juin 1989.
- PI 482 **DEFINITION DE ALPHA : UN LANGAGE POUR LA PROGRAMMATION
SYSTOLIQUE**
Christophe MAURAS
18 Pages, Juin 1989.

